SORT 41 (2) July-December 2017, 347-372 DOI: 10.2436/20.8080.02.63

Joint models for longitudinal counts and left-truncated time-to-event data with applications to health insurance

Xavier Piulachs[1], Ramon Alemany[1], Montserrat Guillén[1] and Dimitris Rizopoulos[2]

Abstract

Aging societies have given rise to important challenges in the field of health insurance. Elderly policyholders need to be provided with fair premiums based on their individual health status, whereas insurance companies want to plan for the potential costs of tackling lifetimes above mean expectations. In this article, we focus on a large cohort of policyholders in Barcelona (Spain), aged 65 years and over. A shared-parameter joint model is proposed to analyse the relationship between annual demand for emergency claims and time until death outcomes, which are subject to left truncation. We compare different functional forms of the association between both processes, and, furthermore, we illustrate how the fitted model provides time-dynamic predictions of survival probabilities. The parameter estimation is performed under the Bayesian framework using Markov chain Monte Carlo methods.

MSC: 62J99, 62N01, 62P05.

Keywords: Joint models, panel count data, left truncation, Bayesian framework, health insurance.

1. Introduction and motivation

The developed world is experiencing significant growth in its elderly population, which not only means people are living longer, but that they tend to face a greater number of years affected by a range of health problems. In the context of health insurance, the changing demographic structure of the population leads to a steady rise in demand for medical services, while the increasing usage of health care systems, in turn, extends longevity even further. This is especially true of private health insurance policyholders, as they are assumed to enjoy greater preventive care than the rest of the population (see e.g., Dow et al., 2010; Chen et al., 2012). Given these circumstances, assessing the

[1] *Corresponding author:* xavier.piulachs@ub.edu. Department of Econometrics, Riskcenter-IREA, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain.

[2] Department of Biostatistics, Erasmus University Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands.

Received: May 2016

Accepted: July 2017

relationship between subject-specific medical history and time until death is of obvious interest for elderly policyholders, as they seek fair premiums. Likewise, insurance companies share this interest, as the must determine the potential costs associated with people living longer than mean expectations. Building on such a scheme, joint modelling techniques are postulated as a proper way to relate the historical information on medical records and the time-to-event outcomes.

The research is conducted on a real health insurance dataset of insured subjects aged 65 years and over, a cohort that requires critical medical care more frequently than their younger counterparts, and, consequently, they have more difficulties in finding private coverage at a reasonable price. Our data contain information both on their health care use and lifespan, and we aim to explain, at subject-level, the underlying mortality risk using the relationship between emergency medical services demanded and time until death. Specifically, the variable of interest in the longitudinal part is the annual rate of emergency claims, including ambulance services, hospitalizations, and non-routine visits. The data only consider the subjects who reach the age of 65, defined in the study as the pre-specified time zero. This assumption has two practical consequences: a) those subjects who die before 65 years of age are not observed, and therefore their time-to-event outcomes are not included, and b) all subjects entering the study after the age of 65 are considered as delayed entries, so their time-to-event data are left-truncated further than the usual censorship (Uzunogullari and Wang, 1992; Klein and Moeschberger, 2003), and not all subjects present the same number of longitudinal measurements. In order to avoid an overestimation of the survival probabilities, a proper consideration of the left truncation issue in the mortality risk is achieved by using the subject's age above 65 years as the particular time scale (Lamarca et al., 1998; Thiébaut and Bénichou, 2004).

The relationship between longitudinal and time-to-event processes can be properly analysed using a shared-parameter joint model (JM), where the corresponding outcomes are stochastically correlated by means of a common latent structure. Using this approach, longitudinal and event times are independent given the random effects, as are repeated measurements in the longitudinal process. Complete overviews of the joint modelling techniques can be found in Tsiatis and Davidian (2004) and Yu, Taylor and Sandler (2008). An exhaustive explanation of the shared-parameter JM, with different examples, is provided by Rizopoulos (2012). In the context of the application of joint modelling techniques to health insurance studies, previous work can be found in Piulachs et al. (2015), where the study focused on elderly policyholders and the counting process was approximated by a log-transformation of the longitudinal outcome.

Given the discrete nature of emergency claims per year, the longitudinal response must account for non-Gaussian data. Previous approaches of this kind have been proposed. For example, Rizopoulos and Ghosh (2011) defined a Bayesian JM to relate multiple longitudinal outcomes (discrete or continuous) to a time-to-event outcome. Murawska, Rizopoulos and Lessaffre (2012) presented a two-stage JM where the longitudinal information was summarized by either a non-linear mixed-effects model or a generalized linear mixed model (GLMM) in the first stage, while in the second the Em-

pirical Bayes estimates of the subject-specific parameters were included as predictors in the proportional hazards model. Viviani, Alfó and Rizopoulos (2012) implemented an expectation-maximization algorithm to incorporate non-Gaussian data in the longitudinal response, with particular attention to Poisson and binomial mixed models. More recently, Ivanova, Molenberghs and Verbeke (2016) formulated a JM to handle different types of responses, i.e., continuous, discrete and ordinal. Parameters were estimated under a likelihood-based approach.

A common feature of the aforementioned extensions is that they do not account for delayed entries in the time-to-event sub-model. In contrast, we consider here the lifetime elapsed from the moment a subject is 65 until his or her death. As a consequence, left truncation has to be accounted for in survival times of these subjects entering the study above the age of 65. Additionally, most event times cease to be observed at administrative closure of study, whereas some others are not completely observed due to dropout. In order to simultaneously deal with left-truncated and right-censored event times, a Cox proportional hazards model with time-dependent covariates is used for the survival analysis. Our final goal is to assess, in a personalized manner, the relationship between emergency claims per year and the time until death (i.e. subject's mortality risk) by postulating an appropriate JM. In this regard, we investigate the role played by information contained in medical records and identify a cumulative and fading effect, so that more recent records have a greater influence than older records on the hazard of death. Finally, we illustrate how the fitted JM can also be employed to obtain subject-specific survival estimates. From a statistical perspective, this problem requires an innovative application of a joint framework, where a pronounced dependency pattern between longitudinal and time-to-event outcomes for the elderly is expected. From a methodological perspective, the statistical analysis poses challenges in handling correlated counts in the longitudinal response of the JM, and to incorporate the delayed entries in the survival outcome.

The remainder of the paper is organized as follows. Section 2 includes a description of the study's health insurance dataset, which consists of 5470 policyholders aged 65 years and over. Section 3 presents the specification, under the Bayesian framework, of the proposed JM for longitudinal counts and left-truncated time-to-event outcomes. Section 4 shows the application of the derived JM to our health insurance dataset, and the results are commented. Section 5 illustrates how to obtain personalized and time-dynamic predictions for survival from the fitted JM. Finally, Section 6 presents a final discussion and some concluding remarks.

2. Health insurance dataset

The motivating dataset was provided by a Spanish medical insurance company, and consists of a cohort of 5470 policyholders (37.6% men and 62.4% women), aged 65 years and above, living in the city of Barcelona (Spain). The data contain, for each subject, historical information on emergency claims (use of ambulance services, hospitaliza-

tions, and non-routine visits) from January 1, 2006 to February 1, 2014. We also know the age of each subject upon entry into the study and their age at death or right censoring, where the latter is assumed to be independent of all other survival and covariate information.

A set of control points was fixed at the 31st of December each year throughout the study period, and we collected, for each subject, subsequent measurements of the amount of emergency claims demanded in a calendar year; this time unit is the one used by most insurance companies, and, in general, in actuarial studies. Hence, instead of directly working with the amount of observed counts, the main longitudinal outcome in our study is defined in terms of count rates. In our case, for each subject we observed repeated measurements of emergency claims per year. These measurements were recorded at each of the control points covered by each of the subject-profiles. In this regard, the entry of each subject into the study period was registered in their longitudinal response by the measurement associated with the first control point reached by his or her observed profile. We assumed a *last observation carried forward* approach for handling the longitudinal information, i.e. an observed measurement within each subject's profile remains constant between two subsequent control points. However, not all subject-profiles started to be observed at the beginning of a specific calendar year. This resulted in the first measurement of emergency claims having an exposure time less than one year. We therefore needed to explicitly consider exposure effects in order to avoid spurious effect estimates (Cameron and Trivedi, 1998). This procedure was carried out by relating the amount of emergency claims observed at the end of a calendar year to the corresponding exposure, i.e. counts/exposure, thus taking into account the real period-at-risk in which the aforementioned amount was collected. Since this premise assumes that the likelihood of a emergency claim is constant over time, very large (and therefore unrealistic) values of count rates could be obtained in case of very small exposures. These cases were avoided by imposing a rule that each of the values registered for a subject must have been obtained from an exposure above half a year.

Table 1: *Descriptive statistics of observed emergency claims per year stratified by event indicator.*

Death	Subjects	*Emergency claims per year* summary				
		Mean	SD	Min	Max	% Zeros
No	4961	0.80	1.55	0	20	63.8
Yes	509	1.50	2.45	0	18	52.4
Overall	5470	0.84	1.63	0	20	63.1

The longitudinal outcome across all count rates ranges from 0 to 20 emergency claims per year, and the overall mean and variance values are 0.84 claims/year and 2.66 $(\text{claims/year})^2$, respectively, suggesting a marked heterogeneity in the response (see Table 1). A large number of zeros are exhibited in the longitudinal outcome, representing 63.1% of the overall measurements. Here, it must be pointed out that the Spanish health system offers universal coverage, so a rate count of zero may occur either because insur-

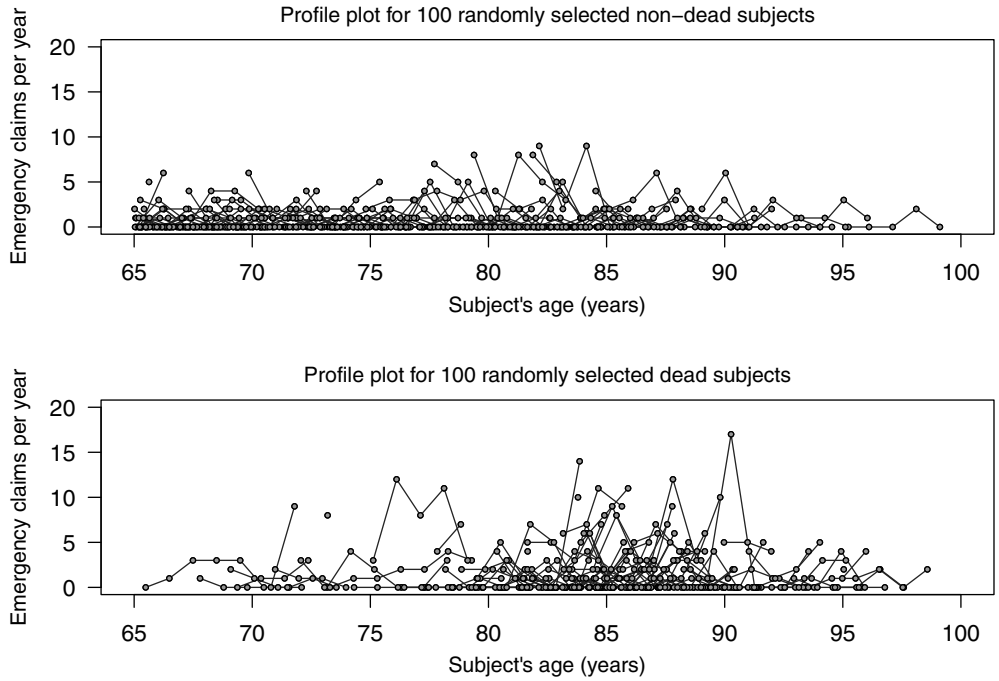

Figure 1: *Subject-profiles of emergency claims per year across time (subject's age) for 100 randomly selected subjects who are still alive after their follow-up interval (top panel) and for 100 randomly selected subjects whose death is observed (bottom panel).*

ance coverage is used solely for routine medical care, or due to the fact that policyholders have only been treated in public medical centers. This circumstance is, in general, an important source of overdispersion in the longitudinal response.

Figure 1 shows various subject-profiles of emergency claims per year, where measurements collected for each subject are connected by line segments. The top panel shows the trajectories for a random sample of 100 subjects alive after their follow-up interval, while the bottom panel shows 100 randomly selected profiles of subjects whose death event is recorded during the study period. Notice that the group of subjects who died during the study presents, in average terms, higher longitudinal responses than those presented by the subjects who remain alive.

Following the suggestions of Charpentier (2015), we also analysed the evolution of the average demand for emergency claims per year according to policyholder's age. We fitted the average values by a generalized additive model (GAM) under the Poisson (PO) and negative binomial (NB) distributions (see Figure 2), and a changing trend was detected around the age of 90 years. Thus, our data show that the use of emergency services in the health insurance company decreases among those subjects of an advanced age. This may reflect the fact that a fraction of the elderly population have taken up residence in nursing homes at older ages, and thus, receive personalized care, or it might be a result of a preference for public over private treatment for severe conditions.

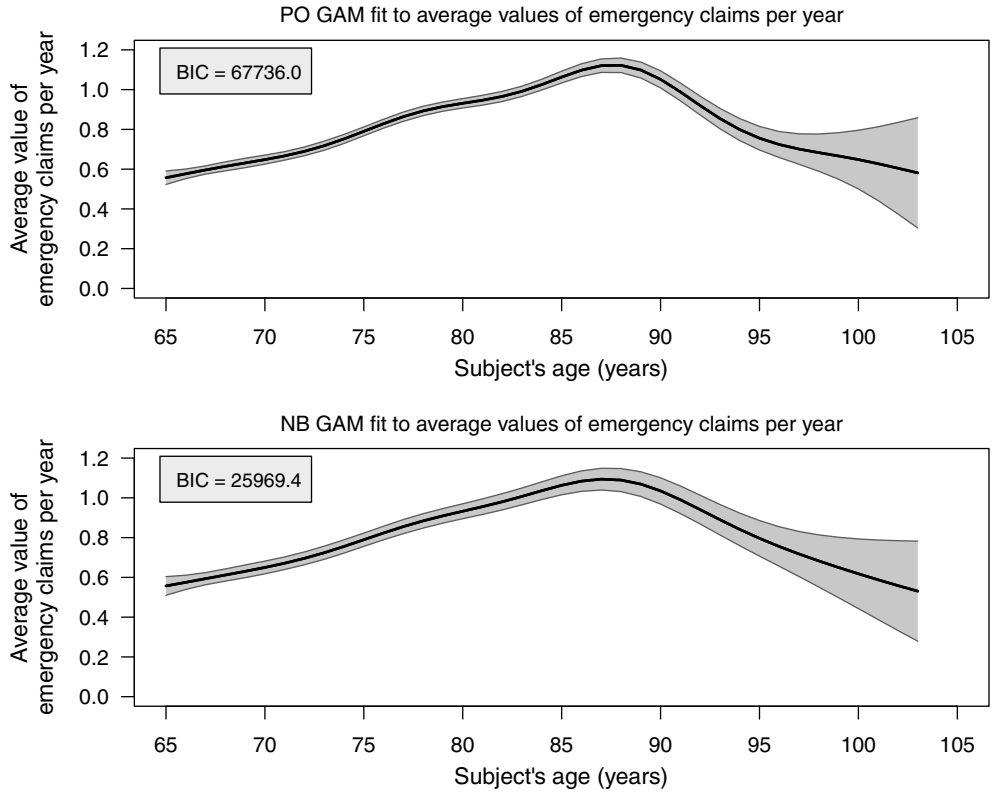

Figure 2: *Observed annual rates of emergency claims by age, with PO and NB GAM fittings. The 95% confidence regions are presented.*

Only policyholders living above the age of 65 are considered within the study period 2006-2014, which means that 79.8% of subjects are registered as late entries. The mean age of policyholders entering the study is 75.4 years (i.e., 10.4 years above the pre-specified time zero), with an average follow-up of 5.1 years. Furthermore, a classic right censoring mechanism arises, which is assumed to be independent of all other survival information. During the study period, death is recorded for a total of 509 (9.3%) individuals, entailing that 4961 policyholders survive or are no longer under observation at the end of the study, representing 90.7% of right censoring. Of these, 3429 (69.1%) are alive at the administrative closure of study, on February 1, 2014. The remaining 1532 right-censored survival times (30.9%) are attributable to insurance cancellations caused by different reasons not related to the event of death (e.g., dissatisfaction with the medical services, a change of company, or an unwillingness to pay), which in practice means that the subject is no longer covered by the insurance policy. Figure 3 displays a non-parametric survival curve estimate of the overall sample (on the left) and one stratified by gender (on the right). Although higher survival estimates are registered for women, the corresponding log-rank test does not suggest a significant improvement in women's survival when stratifying by gender ($p = 0.242$).

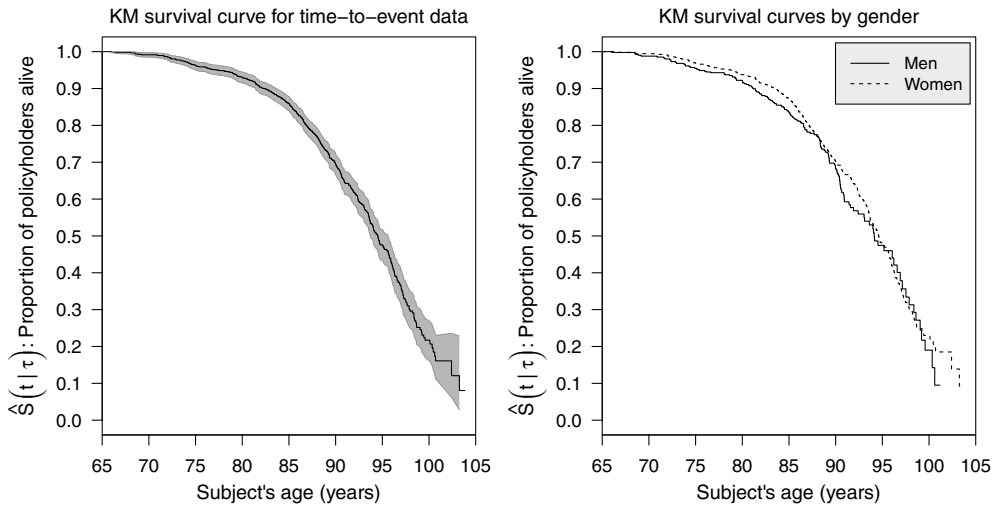

Figure 3: *Kaplan-Meier estimate of the survival function of time until death (with 95% confidence intervals) for our overall health insurance dataset (left panel), and stratified by gender (right panel).*

3. Joint model specification

3.1. Longitudinal approach to panel count data

Let us assume a panel data context with repeated measurements over time, where $\boldsymbol{y}_i = \{y_i(t),\ i = 1,\ldots,n\}$ denote the observed responses for the i-th subject, recorded at a fixed set of time points $t_{ij},\ j = 1,\ldots,n_i$. Given the vector $\boldsymbol{b}_i$ of random effects for the i-th subject, we assume that the observed measurements on this individual derive from a counting process generated by an exponential family (EF) distribution, $y_i(t) \,|\, \boldsymbol{b}_i \sim \text{EF}\{\psi_i(t), \phi\}$, with probability mass function:

$$p_y\{y_i(t) \,|\, \boldsymbol{b}_i; \psi_i(t), \phi\} = \exp\Big(\phi^{-1}\big[y_i(t)\,\psi_i(t) - b\{\psi_i(t)\}\big] + c\{y_i(t), \phi\}\Big). \qquad (1)$$

Here, $b(\cdot)$ and $c(\cdot)$ are known functions, and $\psi_i(t)$ and ϕ are termed the canonical and scale parameters, respectively. It can be shown straightforwardly that $\mathbb{E}\{y_i(t) \,|\, \boldsymbol{b}_i\} = \mu_i(t) = b'\{\psi_i(t)\}$ and $\mathbb{V}\{y_i(t) \,|\, \boldsymbol{b}_i\} = \sigma_i^2(t) = \phi\, b''\{\psi_i(t)\}$ (Molenberghs and Verbeke, 2005).

In many longitudinal studies, the subject-specific count responses are observed within a pre-specified time interval, and can be implicitly interpreted as frequency rates. In such cases, modelling the count rates is more relevant than working with the raw counts, thus considering the expected longitudinal outcome $\mu_i(t)$ in terms of counts per time unit. In our case, a calendar year is taken as the reference time frame during which emergency claims uniformly occur, but a small percentage of subjects start to be observed after the

beginning of a calendar year (i.e., their first longitudinal measurement is not recorded for the duration of a whole year). With this data pattern, the set of observations for a specific subject at their corresponding control points might have occurred during different lengths of time, making it necessary to explicitly consider exposure effects. For the i-th subject at time t, an exposure term $e_i(t)$ is included as a predictor variable of the expected longitudinal outcome $\mu_i(t)$. In addition, it is necessary to introduce a continuous and differentiable link function $g(\cdot)$ in order to relate $\mu_i(t)$ to a linear combination $\eta_i(t)$ of a set of fixed and random covariates. The most common choice for modelling panel count rates is a logarithmic link, $g(\cdot) = \log(\cdot) \Rightarrow g^{-1}(\cdot) = \exp(\cdot)$, which ensures positive outcomes and provides a straightforward interpretation of the estimated regression parameters:

$$\begin{cases} \log\{\mu_i(t)\} = \log\{e_i(t)\} + \eta_i(t) = \log\{e_i(t)\} + \boldsymbol{x}_i^{\intercal}(t)\boldsymbol{\beta} + \boldsymbol{z}_i^{\intercal}(t)\boldsymbol{b}_i \\ \mathbb{E}\{y_i(t) \,|\, \boldsymbol{b}_i\} = \mu_i(t) = e_i(t)\exp\{\eta_i(t)\}, \;\; \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}_{q+1}). \end{cases} \tag{2}$$

Note in the above equation that the exposure term is logged and included as an offset variable, i.e., a predictor whose coefficient is fixed at one. If we move the exposure to the left side of the equation, we evince the fact that our expected outcome is divided by the length of time, $\mu_i(t)/e_i(t)$, so we are effectively modelling the expected response in terms of rate counts. The terms $\boldsymbol{x}_i^{\intercal}(t)$ and $\boldsymbol{z}_i^{\intercal}(t)$ denote the row vectors of the fixed and random design matrices, respectively, while $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^{\intercal}$ and $\boldsymbol{b}_i = (b_{i0}, b_{i1}, \ldots, b_{iq})^{\intercal}$ are the corresponding fixed-effects and random-effects vectors. The random effects allow for the expression of individual deviations from the overall trend, and in most cases they can be assumed to follow a multivariate normal distribution with zero mean and unspecified variance-covariance matrix $\boldsymbol{D}_{q+1}$.

The basic option for modelling panel counts in equation (2) is to consider a PO mixed model, defined as:

$$\begin{cases} y_i(t) \,|\, \boldsymbol{b}_i \sim \text{PO}\{\mu_i(t)\}, \;\; \mu_i(t) > 0 \\ \mu_i(t) = e_i(t)\exp\{\eta_i(t)\} = e_i(t)\exp\{\boldsymbol{x}_i^{\intercal}(t)\boldsymbol{\beta} + \boldsymbol{z}_i^{\intercal}(t)\boldsymbol{b}_i\} \\ p_y\{y_i(t) \,|\, \boldsymbol{b}_i; \mu_i(t)\} = \dfrac{\exp\{-\mu_i(t)\}\,\mu_i(t)^{y_i(t)}}{y_i(t)!} \\ \mathbb{E}\{y_i(t) \,|\, \boldsymbol{b}_i\} = \mathbb{V}\{y_i(t) \,|\, \boldsymbol{b}_i\} = \mu_i(t). \end{cases} \tag{3}$$

The PO mixed model allows for robust parameter estimates, even if the underlying distribution is not true, provided that the expectation is correctly specified (Gourieroux, Monfort and Trognon, 1984). However, the observed response usually has a variance greater than the mean, so the longitudinal outcome is affected by overdispersion. This is a common issue when dealing with counts or count rates, primarily due to missing information, aggregate data, or even an excess of zeros in the longitudinal outcome (Harrison,

2014). In such cases, the derived inference under the PO mixed model could lead to erroneous conclusions about parameter significance. A detailed discussion of this issue can be found in Zuur et al. (2009) and Hilbe (2011).

Although there are several alternative models for dealing with the overdispersion related to correlated counts, the NB mixed model appears in the literature as being the most natural choice; see, for example, Ismail and Jemain (2007), Greene (2008), and Hilbe (2011). The NB distribution for longitudinal data can be easily derived from the PO distribution by placing a multiplicative gamma random noise ε_i in the conditional mean response. Specifically, such a latent variable is defined in terms of shape and rate parameters by $\varepsilon_i \sim G(\kappa,\kappa)$, $\kappa > 0$, with $\mathbb{E}(\varepsilon_i) = 1$ and $\mathbb{V}(\varepsilon_i) = 1/\kappa$, so that the longitudinal counts are modelled by $y_i(t)\,|\,\boldsymbol{b}_i \sim \text{PO}\{\varepsilon_i\,\mu_i(t)\}$. This Poisson-gamma mixture has a closed-form solution, leading to a NB mixed model with dispersion parameter κ:

$$\begin{cases} y_i(t)\,|\,\boldsymbol{b}_i \sim \text{NB}\{\mu_i(t),\kappa\},\ \mu_i(t) > 0, \kappa > 0 \\ \mu_i(t) = e_i(t)\exp\{\eta_i(t)\} = e_i(t)\exp\{\boldsymbol{x}_i^{\intercal}(t)\boldsymbol{\beta} + \boldsymbol{z}_i^{\intercal}(t)\boldsymbol{b}_i\} \\ p_y\{y_i(t)\,|\,\boldsymbol{b}_i;\mu_i(t),\kappa\} = \dfrac{\Gamma\{\kappa + y_i(t)\}}{\Gamma(\kappa)\,y_i(t)!}\ \dfrac{\mu_i(t)^{y_i(t)}\,\kappa^{\kappa}}{\{\mu_i(t)+\kappa\}^{\kappa+y_i(t)}} \\ \mathbb{E}\{y_i(t)\,|\,\boldsymbol{b}_i\} = \mu_i(t);\ \ \mathbb{V}\{y_i(t)\,|\,\boldsymbol{b}_i\} = \mu_i(t) + \mu_i(t)^2/\kappa, \end{cases} \tag{4}$$

where $\Gamma(\cdot)$ denotes the gamma function.

The NB distribution has the general canonical form of the exponential family equations for any fixed κ. Because of the quadratic expression for the variance, it is sometimes referred to as NB type 2 in the literature. Note that the NB distribution can actually be understood as an extension of the PO distribution when overdispersion is accounted for by parameter κ, since it can be proven that NB converges to PO as $\kappa \to \infty$. This result is well-documented by Lawless (1987) and Hinde and Demétrio (1998); see also Boucher, Denuit and Guillén (2008) for a numerical application in the field of insurance studies.

3.2. Joint model for counts and delayed entries

Assuming the age above 65 years as our particular time scale, let T_i^* be the true event time for the i-th subject. We define an independent random variable $\tau_i \geq 0$ as the time at which a policyholder enters the study after the age of 65, giving rise to left-truncated event times for those subjects whose $\tau_i > 0$. In addition, once a subject enters the study, the event time is affected by the usual right censorship mechanism, denoted by a potential censoring time C_i. This means we can only know the observed survival time for the i-th recruited individual, $T_i = \min\{T_i^*, C_i\} > \tau_i$, and a dichotomous event indicator $\delta_i = \mathbb{I}(T_i^* \leq C_i)$. We use a time-dependent proportional hazards model to simultaneously account for left truncation and right censoring in the time-to-event sub-model. Conse-

quently, the probabilistic distribution of the event times has to be defined according to the proportion of subjects living beyond time point t, and is conditional on their being older than the corresponding left truncation time, $S(t \mid \tau) = \Pr(T^* > t \mid T^* > \tau) = \Pr(T^* > t) \,/\, \Pr(T^* > \tau) = S(t) \,/\, S(\tau)$.

Building on the longitudinal analysis considered in section 3.1, repeated count rates and time-to-event responses can be coupled by assuming independence between both processes given the shared random effects (conditional independence hypothesis). The JM for the i-th subject, $i = 1, \ldots, n$, is expressed by means of a relative risk model where the hazard of death at time t takes into account the expected longitudinal response until t, $M_i(t) = \{\mu_i(s), 0 \leq s \leq t\}$:

$$h_i\{t \mid M_i(t), \boldsymbol{w}_i\} = h_0(t) \exp\left[\boldsymbol{\gamma}^{\mathsf{T}} \boldsymbol{w}_i + \alpha F\{\mu_i(t)\}\right]. \tag{5}$$

As in the standard proportional hazards model, $h_0(t)$ in equation (5) denotes the baseline risk function, $\boldsymbol{w}_i$ the subject's baseline survival covariates, and $\boldsymbol{\gamma}$ the vector of the corresponding regression parameters. The functional form $F(\cdot)$ specifies a proper manner in which the longitudinal information provided by $\mu_i(t)$ is accounted for in survival. Because $\mu_i(t) > 0$ in a counting process, $F(\cdot)$ is positively defined and increases with t. The parameter α quantifies the strength of association between the particular longitudinal evolution until time t, and the corresponding mortality risk. Specifically, the quantity $\exp(\alpha)$ returns the hazard ratio for a one-unit increase in the value $F\{\mu_i(t)\}$ at time snapshot t.

Although $h_0(t)$ traditionally remains unspecified in the Cox proportional hazards model, this constraint is usually lifted when using joint modelling techniques. In particular, the logarithm of baseline hazard function can be approximated using penalized B-splines. As a preliminary step, we define a knot sequence $\boldsymbol{\xi}$ of Q increasing and equally-spaced knots, $\xi_1 < \cdots < \xi_Q$, over the time range $[0, T_{max}]$. Accordingly, the baseline hazard on the log-scale is approximated through a linear combination of d-th degree B-splines:

$$\log\{h_0(t)\} = \sum_{r=1}^{R} \gamma_{h_0,r} B_{d,r}(t, \boldsymbol{\xi}), \tag{6}$$

where $\{B_{d,r}(t, \boldsymbol{\xi}),\ r = 1, \ldots, R\}$ denotes the set of d-th degree B-spline basis functions, $\boldsymbol{\gamma}_{h_0} = (\gamma_{h_0,1}, \ldots, \gamma_{h_0,R})^{\mathsf{T}}$ is the vector of B-spline coefficients (also called control points), and $R = Q + d - 1$. The r-th B-spline function is locally defined on a support spanned by the $d+2$ adjacent knots, and, to achieve boundary conditions of a B-spline curve, the original knot vector is extended so that the end-knots ξ_1 and ξ_Q have multiplicity $d+1$ (the total number of knots will be $Q + 2d$). A major concern at this point is the number Q of knots that should be employed. A too-small number of knots (and, consequently, of the number R of B-spline basis functions) could lead to biased results, while too many knots might result in an overly flexible curve with random fluctuations (small "wig-

gles"). Following the indications of Eilers and Marx (1996), a relatively large number of knots should be used, and the potential overfitting problems can be circumvented by considering a roughness penalty based on finite differences of adjacent B-spline coefficients, i.e., by means of a P-splines regression. A complete overview of recent research in P-splines can be found in Eilers, Marx and Durbán (2015).

A standard approach to relate longitudinal rate counts to survival is undertaken by associating the current expected longitudinal outcome with the hazard of an event using the identity function: $F\{\mu_i(t)\} = Id\{\mu_i(t)\} = \mu_i(t)$. However, instead of taking just a single time point, in some cases it may be more relevant to consider the whole path of the longitudinal outcome. In particular, an extension of the basic option is to include the entire background previous to the measurement at time t (Abrahamowicz, Beauchamp and Sylvestre, 2011). Furthermore, we assume that historical effects of subject's health fade over time, so the more distant history is less relevant than the more recent. Thus, $F(\cdot)$ transformation can be defined to account for the recency-weighted cumulative effect of the longitudinal outcome:

$$F\{\mu_i(t)\} = \int_0^t \overline{\omega}(t-s)\,\mu_i(s)\,\mathrm{d}s, \quad s \leq t, \tag{7}$$

where $\overline{\omega}(\cdot)$ is the selected average weighting function. Due to the importance of the most recent information for explaining the current health status, we introduced an exponential function with rate parameter ν in order to assign different weights for each of the past observed longitudinal values: $\overline{\omega}(t-s) = \nu \exp\{-\nu(t-s)\},\ \nu > 0$.

3.3. Bayesian estimation for the JM

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\theta}_b)^{\mathsf{T}}$ be the JM full parameter vector that collects the longitudinal parameters, the survival parameters, and the parameters for the random effects covariance matrix, respectively. In addition, let $D_n = \{(\boldsymbol{y}_i, \tau_i, T_i, \delta_i),\ i = 1, \ldots, n\}$ denote the information from our original dataset with n policyholders. Taking advantage of the conditional independence assumption, the overall joint likelihood conditioned on the random effects $\boldsymbol{b}_i$ can be properly formulated to tackle left truncation as

$$p(D_n \,|\, \boldsymbol{b}_i, \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{n_i} p_y\{y_i(t_{ij}) \,|\, \boldsymbol{b}_i, \boldsymbol{\theta}\}\, \frac{p_t(T_i, \delta_i \,|\, \boldsymbol{b}_i, \boldsymbol{\theta})}{\Pr(T_i > \tau_i \,|\, \boldsymbol{b}_i, \boldsymbol{\theta})}, \tag{8}$$

where $p_y(\cdot)$ is the conditional probability mass function to handle longitudinal rate counts, and $p_t(\cdot)$ is the conditional probability density function for the event times.

The mean estimates of parameters and random effects are then derived by Markov chain Monte Carlo (MCMC) algorithms, which enable inferences to be made by efficiently drawing a sample from the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{b}_i)$ conditioned on the observed data:

$$\pi(\boldsymbol{\theta}, \boldsymbol{b}_i \,|\, D_n) \;\propto\; p(D_n \,|\, \boldsymbol{b}_i, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}, \boldsymbol{b}_i) = p(D_n \,|\, \boldsymbol{b}_i, \boldsymbol{\theta})\, p_b(\boldsymbol{b}_i \,|\, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta}), \tag{9}$$

where $p_b(\cdot)$ is the conditional probability density function of the random effects, and $\pi(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$.

The models' fitting was performed using a Bayesian approach, with non-informative priors being used whenever possible. Specifically, for the longitudinal analysis, we used independent univariate vague normal priors for the fixed effects, defined in terms of mean and precision parameters, $\{\beta_0, \beta_1, \beta_{t_{90}}\} \sim N\left(\mu_\beta = 0, \tau_\beta = 10^{-3}\right)$. In the time-to-event sub-model, the log-baseline hazard was approximated by using third-degree B-splines and $Q = 15$ knots, uniformly allocated over the time range $[0,\, T_{max} = 38.5]$. The joint prior for the baseline hazard coefficients was assumed to be normally distributed, $\boldsymbol{\gamma}_{h_0} \sim N(\boldsymbol{\mu}_{\boldsymbol{\gamma}_{h_0}} = \boldsymbol{0},\, \boldsymbol{T}_{\boldsymbol{\gamma}_{h_0}} = \tau_{Bs}\boldsymbol{M}_{\boldsymbol{\gamma}_{h_0}})$, where $\boldsymbol{M}_{\boldsymbol{\gamma}_{h_0}}$ is an appropriate penalty matrix to control the amount of roughness in the precision parameter $\tau_{Bs} \sim G(a_{Bs}, b_{Bs})$. In general, the penalty matrix is defined as $\boldsymbol{M}_{\boldsymbol{\gamma}_{h_0}} = \boldsymbol{\Delta}_K^{\intercal}\boldsymbol{\Delta}_K + 10^{-6}\boldsymbol{I}$, where $\boldsymbol{\Delta}_K$ is the difference matrix of order K, whereas the term $10^{-6}\boldsymbol{I}$ introduces a small "ridge penalty" to avoid a linearly dependent system. A common choice for cubic B-splines is $K = 2$, while for the hyper-prior parameters of τ_{Bs} we used $a_{Bs} = 1$ and $b_{Bs} = 0.005$ as a standard choice for a non-informative prior. The subject's gender was included in the time-dependent proportional hazards model as a dichotomous baseline covariate, w_{gi} (man $= 0$, woman $= 1$), and we made the assumption that the corresponding coefficient follows an improper normal distribution, $\gamma_g \sim N\left(\mu_g = 0, \tau_g = 10^{-3}\right)$.

For the constant association parameter, we assumed $\alpha \sim N\left(\mu_\alpha = 0, \tau_\alpha = 10^{-3}\right)$. When considering the functional form to link the recency-weighted area under the expected longitudinal profile to the time-to-event outcome, a flat prior was assumed for the rate parameter of the exponential weighting function, $\nu \sim U(a_\nu, b_\nu)$. Because $\nu > 0$, we set $a_\nu = 0$, while for the second hyper-parameter it is common to set a large enough positive value to express the uncertainty around ν, say $b_\nu = 20$.

Finally, for the random effects, we used a bivariate standard normal distribution as a prior function, $(b_{i0},\, b_{i1})^{\intercal} \sim N(\boldsymbol{0}, \boldsymbol{D}_2)$, where the terms of the 2×2 unstructured covariance matrix are summarized by $\boldsymbol{D}_2[1,1] = \sigma^2_{b_0}$, $\boldsymbol{D}_2[1,2] = \boldsymbol{D}_2[2,1] = \rho\,\sigma_{b_0}\,\sigma_{b_1}$, and $\boldsymbol{D}_2[2,2] = \sigma^2_{b_1}$. We assumed that the inverse matrix follows a standard Wishart distribution, $\boldsymbol{D}_2^{-1} \sim W(\boldsymbol{I}_2, k_w)$, where the degrees of freedom were established at $k_w = 3$. In the particular case of considering 1 RE, we have $b_{i1} = \sigma_{b_1} = 0$, assuming that $b_{i0} \sim N(\mu_{b_0}, \tau_{b_0})$ with $\mu_{b_0} = 0$ and $\tau_{b_0} \sim G\left(10^{-3}, 10^{-3}\right)$.

3.4. Bayesian model assessment

To compare both the different longitudinal models and joint models, we focused on the analysis of the Bayesian deviance term, which in generic form can be expressed as $\mathrm{D}(\boldsymbol{\theta}, \boldsymbol{b}_i) = -2\sum_{i=1}^{n} \log\{p\,(D_n \,|\, \boldsymbol{b}_i, \boldsymbol{\theta})\}$. In particular, we assessed the goodness-of-fit of a specific model by using the deviance information criterion (DIC) suggested by

Spiegelhalter et al. (2002). This criterion evaluates the fit of a model by balancing model adequacy with model complexity:

$$\mathrm{DIC}(\boldsymbol{\theta}, \boldsymbol{b}_i) = \mathrm{D}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{b}}_i) + 2\,\mathrm{p_D}, \tag{10}$$

where $\mathrm{D}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{b}}_i) = -2\sum_{i=1}^{n} \log\{p\,(D_n \,|\, \bar{\boldsymbol{b}}_i, \bar{\boldsymbol{\theta}})\}$, and the term $\mathrm{p_D} = \overline{\mathrm{D}(\boldsymbol{\theta}, \boldsymbol{b}_i)} - \mathrm{D}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{b}}_i)$ is the effective number of parameters, calculated as the difference between the posterior mean of the deviance and the deviance at the posterior means of the JM parameters. The aforementioned criterion can be reformulated as $\mathrm{DIC}(\boldsymbol{\theta}, \boldsymbol{b}_i) = \overline{\mathrm{D}(\boldsymbol{\theta}, \boldsymbol{b}_i)} + \mathrm{p_D}$, thus reinforcing the idea that this criterion takes into account both the adequacy of the model, assessed through the posterior mean estimate of the deviance, and the number of parameters required, assessed through the penalty term $\mathrm{p_D}$. The score provided by DIC serves in general as the basis for ranking the fitted models, where lower scores correspond to a better model fit. To conclude this section, it is important to point out that the DIC score obtained for a specific model is not a fixed value, but it can be subject to a certain amount of random variability due to its dependency on the MCMC output of the model. Consequently, it will become a key point to get a DIC value derived from a relatively large number of iterations in the MCMC process before reaching convergence in each of the JM parameters.

4. Results from the health insurance dataset

4.1. Longitudinal data analysis

The fixed effects of the longitudinal outcome were set at $\{\beta_0, \beta_1, \beta_{t_{90}}\}$, respectively motivated by the intercept term, the observation time (directly linked to the subject's age), and a binary variable which takes into account the observed downward trend of medical emergency demand after the age of 90, $t_{90} = \mathbb{I}(t \geq 25)$. The longitudinal measurements were fitted using PO and NB mixed models, testing these models with random intercepts (1 RE) and with ρ-correlated random intercepts and random slopes in time (2 RE). In all cases, the general expression for the expected response can be written as $\mu_i(t) = e_i(t)\exp\{\eta_i(t)\} = e_i(t)\exp\{\beta_0 + b_{i0} + (\beta_1 + b_{i1})\,t + \beta_{t_{90}}\mathbb{I}(t \geq 25)\}$, with $(b_{i0},\, b_{i1})^{\mathsf{T}} \sim N(\mathbf{0}, \boldsymbol{D}_2)$.

Setting $\boldsymbol{\theta}_\ell = (\boldsymbol{\theta}_y, \boldsymbol{\theta}_b)^{\mathsf{T}} = \left(\beta_0, \beta_1, \beta_{t_{90}}, \kappa, \sigma_{b_0}, \sigma_{b_1}, \rho\right)^{\mathsf{T}}$, we approximated the posterior distribution $\pi(\boldsymbol{\theta}_\ell, \boldsymbol{b}_i \,|\, \boldsymbol{y}_i)$ by using the Bayesian software JAGS (Plummer, 2003), version 4.2.0. We ran two parallel chains of 25,000 iterations, with a burn-in period (here included in the total number of iterations) of 10,000 iterations. We then kept every 15-th iteration from each chain in order to reduce the autocorrelation in the samples from the posterior distribution, resulting in 2000 total samples. We checked that the chains had a good mixing, and also their convergence to the stationary distribution. For

each model, the DIC was calculated by means of the corresponding Bayesian deviance $\mathrm{D}(\boldsymbol{\theta}_\ell, \boldsymbol{b}_i) = -2\sum_{i=1}^{n} \log p(\boldsymbol{y}_i \,|\, \boldsymbol{b}_i, \boldsymbol{\theta}_\ell)$.

Table 2 provides the posterior summaries of the parameters' distribution for each longitudinal model fitted, and also the corresponding values of DIC. In the case of the two PO mixed models, the DIC provides a lower score for the option which considers 2 RE in the linear predictor. On the other hand, the NB mixed model with 2 RE also provides evidence of an improvement in the DIC score in comparison with the model using 1 RE. Therefore, it is preferable to use a random intercept and a random slope to summarize subjects' profiles. Along with the PO results, the estimates obtained from the NB mixed model with 2 RE highlight the importance of the variability around the intercept fixed effect compared to fluctuations in the slope. Hence, accounting for baseline heterogeneity indeed plays a much more important role to explain the subject's particularities. This could lead us to consider the extent to which the introduction of a random slope is necessary in this case (i.e., when a linear trend is assumed in the specification of $\eta_i(t)$).

Table 2: *Posterior summaries of all parameters for PO and NB mixed models with a different number of random effects. Mean, standard error, 95% credible interval, and DIC are sampled for each parameter from the corresponding posterior distribution.*

	PO mixed model 1 RE				NB mixed model 1 RE			
Parameter	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$
β_0	−1.085	0.0007	−1.147	−1.026	−0.976	0.0015	−1.041	−0.910
β_1	0.032	0.0001	0.029	0.036	0.031	0.0001	0.027	0.035
$\beta_{t_{90}}$	−0.117	0.0008	−0.193	−0.045	−0.189	0.0023	−0.298	−0.086
κ	–	–	–	–	0.998	0.0004	0.948	1.050
σ_{b_0}	1.075	0.0004	1.045	1.107	0.963	0.0006	0.929	0.996
$\mathrm{DIC}(\boldsymbol{\theta}_\ell, \boldsymbol{b}_i)$	76961				74201			
	PO mixed model 2 RE				NB mixed model 2 RE			
Parameter	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$
β_0	−1.158	0.0009	−1.240	−1.075	−0.974	0.0009	−1.049	−0.900
β_1	0.035	0.0001	0.029	0.040	0.029	0.0001	0.024	0.034
$\beta_{t_{90}}$	−0.190	0.0022	−0.281	−0.103	−0.197	0.0015	−0.328	−0.071
κ	–	–	–	–	1.058	0.0007	1.003	1.120
σ_{b_0}	1.750	0.0011	1.662	1.848	1.157	0.0015	1.074	1.239
σ_{b_1}	0.112	0.0001	0.106	0.118	0.074	0.0001	0.069	0.079
ρ	−0.796	0.0003	−0.818	−0.773	−0.614	0.0007	−0.671	−0.547
$\mathrm{DIC}(\boldsymbol{\theta}_\ell, \boldsymbol{b}_i)$	74993				74095			

The overall comparison between the fitted models suggests that the NB mixed models are more adequate to capture the characteristics of the longitudinal data. This result is unsurprising since the two NB mixed models account for response heterogene-

ity through parameter κ, whose mean estimate exhibits strong evidence for overdispersion for both one and two random effects, $\kappa_{1\,\mathrm{RE}} = 0.998$ (95% CI: 0.948, 1.050) and $\kappa_{2\,\mathrm{RE}} = 1.058$ (95% CI: 1.003, 1.120). In particular, the NB mixed model with 2 RE is the one which presents the lowest DIC score among the tested options. Consequently, we decided to include the effect of random slope when overdispersion is accounted for.

In what follows, the longitudinal approach in our JM framework is carried out using a count mixed model with two random intercepts per subject, one for the intercept and the other for the slope. Additionally, the longitudinal analysis will be carried out using either a PO or a NB mixed model, thus allowing us to directly assess how the goodness-of-fit changes when considering the overdispersion effect.

4.2. JM analysis

The JM that we propose is summarized by:

$$\begin{cases} \mu_i(t) = e_i(t)\exp\left\{\beta_0 + b_{i0} + \beta_1 t + b_{i1} t + \beta_{t_{90}} \mathbb{I}(t \geq 25)\right\} \\ h_i\left\{t \,|\, M_i(t), w_{gi}\right\} = h_0(t)\exp\left[\gamma_g w_{gi} + \alpha F\left\{\mu_i(t)\right\}\right] \\ (b_{i0}, b_{i1})^{\intercal} \sim N(\boldsymbol{0}, \boldsymbol{D}_2). \end{cases} \tag{11}$$

The starting point to carry out the JM fits under a Bayesian approach was the `R` package `JMbayes` (Rizopoulos, 2016), taking advantage of the structure of the function `jointModelBayes`$(\cdot)$. However, the code to fit the different joint models in this article was finally written in `JAGS` software, and executed within the `R`-environment. Setting $\boldsymbol{\theta} = (\boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\theta}_b)^{\intercal} = (\beta_0, \beta_1, \beta_{t_{90}}, \kappa, \nu, \boldsymbol{\gamma}_{h_0}, \gamma_g, \alpha, \sigma_{b_0}, \sigma_{b_1}, \rho)^{\intercal}$, the posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{b}_i \,|\, D_n)$ was approximated by running the MCMC algorithm for 2 parallel chains with a total of 35,000 iterations each, with the first 15,000 discarded as the burn-in period. We kept every 20-th iteration from each chain, resulting in 2000 total samples from the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{b}_i)$. A good mixing and convergence of the 2 chains were assessed, and no autocorrelation was detected in the lag plots.

First, the estimation of JM parameters was conducted by quantifying the degree of association between the current expected value of emergency claims per year at any time t, denoted by $\mu_i(t) = \mathbb{E}\left\{y_i(t) \,|\, \boldsymbol{b}_i\right\}$, and the mortality risk at the same t. The results (given in Table 3) point to a strong association between the annual rate of emergency claims and survival, so each unit increase in the current value of the emergency claims per year involves a $\exp(\overline{\alpha}_{\mathrm{V,PO}}) = 1.47$-fold increase (95% CI: 1.41, 1.54) in the policyholder's mortality risk under the PO longitudinal sub-model, whereas this association parameter leads to a $\exp(\overline{\alpha}_{\mathrm{V,NB}}) = 1.59$-fold increase (95% CI: 1.49, 1.71) if we assume a NB longitudinal sub-model. Thus, we have an increasing relationship between the frequency of use of emergency medical services and the corresponding mortality risk. From a goodness-of-fit perspective, the comparison between the fitted joint models is performed using the DIC, where the use of the NB distribution provides a better fit.

Table 3: *Posterior summaries of all parameters for the JM when accounting for the current value of emergency claims per year. Mean, standard error, 95% credible interval, and DIC are sampled for each parameter from the corresponding posterior distribution.*

	JM with PO sub-model				JM with NB sub-model			
Parameter	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$
Longitudinal								
β_0	−1.174	0.0010	−1.254	−1.085	−1.000	0.0008	−1.072	−0.920
β_1	0.036	0.0001	0.030	0.041	0.031	0.0001	0.026	0.037
$\beta_{t_{90}}$	−0.130	0.0010	−0.212	−0.043	−0.117	0.0018	−0.236	−0.004
κ	–	–	–	–	1.067	0.0004	1.012	1.125
σ_{b_0}	1.780	0.0011	1.685	1.872	1.169	0.0019	1.080	1.247
σ_{b_1}	0.115	0.0001	0.109	0.121	0.076	0.0001	0.070	0.081
ρ	−0.800	0.0003	−0.821	−0.775	−0.611	0.0002	−0.626	−0.595
Survival								
γ_g	−0.287	0.0018	−0.449	−0.132	−0.326	0.0019	−0.483	−0.164
Association								
α	0.387	0.0005	0.342	0.431	0.464	0.0011	0.397	0.534
Goodness-of-fit								
DIC($\boldsymbol{\theta}, \boldsymbol{b}_i$)	93050				86938			

Table 4: *Posterior summaries of all parameters for the JM when accounting for the recency-weighted area under the expected profile of emergency claims per year. Mean, standard error, 95% credible interval, and DIC are sampled for each parameter from the corresponding posterior distribution.*

	JM with PO sub-model				JM with NB sub-model			
Parameter	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$	Mean	SE	$q_{2.5\%}$	$q_{97.5\%}$
Longitudinal								
β_0	−1.166	0.0010	−1.250	−1.089	−1.002	0.0015	−1.068	−0.936
β_1	0.035	0.0001	0.030	0.041	0.031	0.0001	0.026	0.036
$\beta_{t_{90}}$	−0.129	0.0016	−0.212	−0.044	−0.116	0.0020	−0.234	−0.003
κ	–	–	–	–	1.066	0.0005	1.011	1.124
ν	9.572	0.0218	8.154	11.060	9.691	0.0275	8.246	11.218
σ_{b_0}	1.770	0.0016	1.698	1.846	1.175	0.0014	1.096	1.261
σ_{b_1}	0.114	0.0001	1.110	1.119	0.076	0.0002	0.071	0.082
ρ	−0.798	0.0003	−0.820	−0.774	−0.607	0.0003	−0.647	−0.561
Survival								
γ_g	−0.269	0.0019	−0.433	−0.092	−0.298	0.0021	−0.489	−0.124
Association								
α	0.398	0.0005	0.354	0.443	0.480	0.0010	0.422	0.537
Goodness-of-fit								
DIC($\boldsymbol{\theta}, \boldsymbol{b}_i$)	92983				86892			

One of the more interesting features of the JAGS software is its flexibility in choosing the structure association $F(\cdot)$ that captures the relationship between the longitudinal and time-to-event sub-models. The JM estimates in Table 4 were conducted by associating

the recency-weighted area under the expected longitudinal profile with the mortality risk. Specifically, an exponential weighting function was employed, again showing a strong relationship between both processes, so that a $\exp(\overline{\alpha}_{\text{W,PO}}) = 1.49$-fold increase (95% CI: 1.42, 1.56) in the policyholder's mortality risk is inferred for each unit increase in the exponentially-weighted area under the expected profile of emergency claims per year with the PO distribution, and a $\exp(\overline{\alpha}_{\text{W,NB}}) = 1.62$-fold increase (95% CI: 1.53, 1.71) if we assume an underlying NB distribution. The estimated mean rates of the exponential weighting functions for the PO and NB longitudinal outcomes are $\overline{\nu}_{\text{PO}} = 9.57$ (95% CI: 8.15, 11.06) and $\overline{\nu}_{\text{NB}} = 9.69$ (95% CI: 8.25, 11.22), respectively. Thus, in practice, it is shown that only the 0.25 years (i.e., three months) prior to t are strongly related to the current mortality risk. In this regard, note the broad similarity between the association parameters of these results and those obtained in Table 3 for the current-value association structure, thus emphasizing that only the most recent past emergency claims have a real influence on the survival. Once again the DIC indicates that a more accurate claims distribution is achieved under the NB longitudinal sub-model.

Among all the fitted joint models presented in this section, the results indicate that the lower DIC scores are obtained for the functional form which links the recency-weighted area under the expected longitudinal outcome with survival. This becomes an adequate manner to capture the fading effect of emergency medical demand on mortality risk. In particular, the JM which considers a NB longitudinal outcome is the one which provides the lowest DIC score of all, since it includes the overdispersion effect.

4.3. Residual diagnostics and model assessment

After fitting the joint models, it is a primary step to validate all the necessary model assumptions before performing inference. To achieve this validation, we need plots of residuals for each of the two components of the JM, i.e. the longitudinal and the time-to-event sub-models.

For the longitudinal part, the analysis of the plots of residuals is focused on the non-Gaussian mixed models assumed for the joint models with a recency-weighted cumulative effect. We will consider both PO and NB distributions in order to compare their results. However, a direct graphical interpretation of the residuals under these distributions is usually difficult, since the normality and homoscedasticity of the residuals derived from a count model is, in general, not expected. When longitudinal response takes a limited number of low count rates, the scatterplot of the residuals versus the fitted values typically shows a non-homogeneous configuration, the data being grouped on a set of quasi-parallel and curvilinear traces of points according to distinct response values. In such circumstances, it becomes difficult to evaluate the residual plot, even if the model is correctly specified. To solve this limitation, we can obtain continuously distributed residuals by taking advantage of the idea of randomized quantile residuals (Dunn and Smyth, 1996). The underlying idea is based on applying a transformation scale to the original residuals that standardizes them to continuous values between 0

and 1, so that the residuals are obtained by finding the equivalent standard normal deviate for each subject-specific observation in the original data. By delving deep into this work scheme, the quantile residuals can also be directly obtained through a simulation-based approach. In particular, this task has been recently implemented in the R package DHARMa (Hartig, 2017), which standardizes the residuals to uniformly distributed values in the unit interval. As a first step, from JM results we use the fitted longitudinal sub-model $p_y\{y_i(t)\,|\,\boldsymbol{b}_i,\overline{\mu}_i(t),\overline{\kappa}\}$ to simulate a relatively large number M of new longitudinal datasets, $\{\tilde{\boldsymbol{y}}_i^{(m)}\}_{m=1}^{M}$. Then, for a particular subject-specific observation $y_i(t)$ in the original data, we have a set of M simulated values $\{\tilde{y}_i^{(1)}(t),\ldots,\tilde{y}_i^{(M)}(t)\}$, allowing us to obtain its corresponding empirical cumulative distribution function, $\tilde{P}_y\{\tilde{y}_i(t)\,|\,\boldsymbol{b}_i,\overline{\mu}_i(t),\overline{\kappa}\}$. Finally, the quantile residual associated with the original observation $y_i(t)$ is calculated as:

$$r_i^q(t) = \tilde{P}_y\{y_i(t)\,|\,\boldsymbol{b}_i,\overline{\mu}_i(t),\overline{\kappa}\} = \Pr\{\tilde{y}_i^{(m)}(t) \leq y_i(t)\} \in (0,1). \tag{12}$$

Recall that, if the longitudinal model is correctly specified, there will be no difference between the original dataset and the M simulated datasets, so all the values in the empirical cumulative distribution will have the same probability. In such a case, this would lead to a uniform distribution of the residuals, regardless of the longitudinal model employed to fit the data. Once the described process has been repeated for each of the original observations, a residual analysis can easily be carried out, detecting deviations from the uniform distribution, residual dependency on a predictor, or overdispersion.

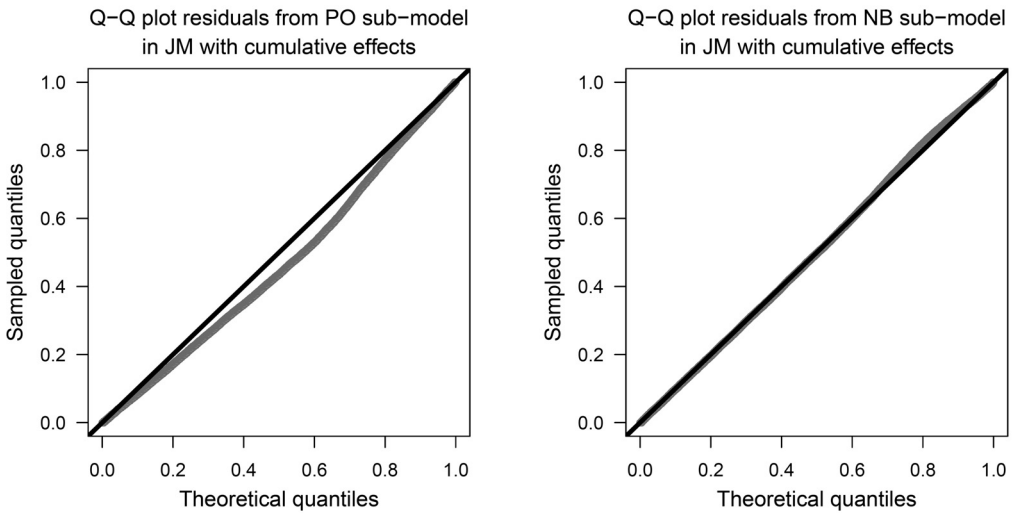

Figure 4: *Randomized quantile residuals for the longitudinal sub-model of the joint models with a recency-weighted cumulative effect, for both PO longitudinal sub-model (left panel) and NB longitudinal sub-model (right panel).*

Figure 4 depicts the randomized quantile residuals derived from both PO and NB longitudinal sub-models in the fitted JM with a recency-weighted cumulative effect parameterization, where $M = 500$ datasets were simulated for each sub-model. The quantile residuals of the estimated PO longitudinal sub-model show clear evidence of lack of fit, exhibiting a quadrature pattern, probably due to the overdispersion effect. In order to confirm this trend, we also plotted the residuals against the time predictor, and again a systematic deviation quadrature was obtained in the corresponding Q-Q plot. By contrast, the quantile residuals of the NB sub-model are almost perfectly uniformly distributed, lying approximately on the diagonal line. These graphical results strongly suggest the adequacy of the NB longitudinal sub-model.

To check the quality of the survival model's predictions, the analysis of martingale residuals (Barlow and Prentice, 1988) is a very common graphical method. Let $R_i(t) = \mathbb{I}(T_i \geq t)$ be the indicator that the i-th subject is at risk at time t, and $N_i(t)$ be the corresponding cumulative number of events until time t. In general, the martingale residuals for subject i at time t is defined by the mean estimates $(\overline{\boldsymbol{\theta}}, \overline{\boldsymbol{b}}_i)$ as

$$r_i^m(t) = N_i(t) - \int_0^t R_i(s)\, h_i\{s \,|\, \overline{M}_i(s), \overline{\boldsymbol{\theta}}\}\, \mathrm{d}s, \quad r_i^m(t) \in (-\infty, 1]. \tag{13}$$

Here, $h_i\{s \,|\, \overline{M}_i(s), \overline{\boldsymbol{\theta}}\} = \overline{h}_0(s) \exp\left[\overline{\gamma}_g w_{gi} + \overline{\alpha} F\{\overline{\mu}_i(s)\}\right]$, where $\overline{h}_0(\cdot)$ is the estimated baseline hazard function of the relative risk model, and $\overline{\mu}_i(s) = e_i(s) \exp\{\overline{\beta}_0 + \overline{b}_{i0} + \overline{\beta}_1 s + \overline{b}_{i1} s + \overline{\beta}_{t_{90}} \mathbb{I}(s \geq 25)\}$. The martingale residuals are calculated, for each subject-specific measurement, as the difference between the observed number of events

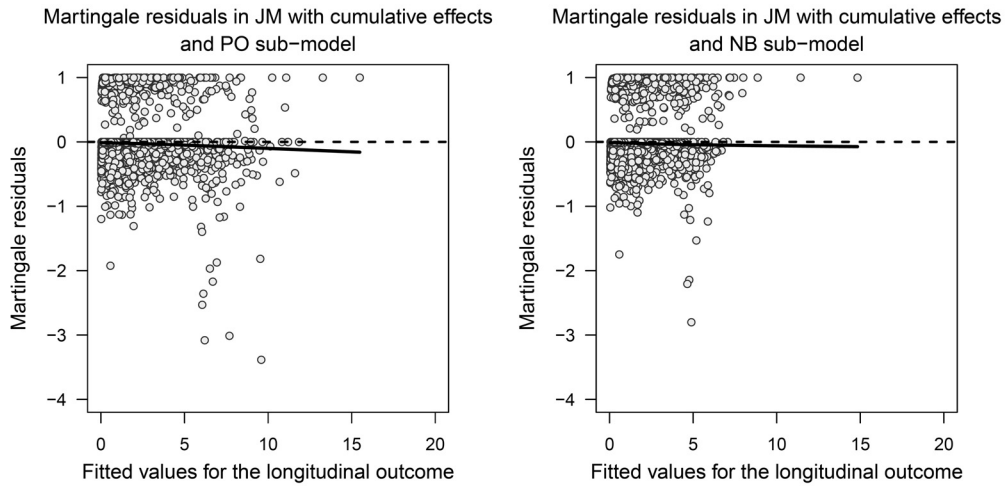

Figure 5: *Martingale residuals derived from the joint models with a recency-weighted cumulative effect. The left panel shows the martingale residuals versus the subject-specific values fitted with the PO longitudinal sub-model, and the right panel shows the results obtained with the NB longitudinal sub-model. In both plots, a loess function has been overlaid to determine the trend.*

(0 or 1) and the number of events expected to happen. This implicitly allows for the identification as outliers of those cases where the survival function predicts an event either too early (values near 1) or too late (extreme negative values). Using an adequate model specification, these residuals should be uncorrelated with one another and have a zero mean, even though they are not symmetrically distributed around zero value.

Taking the results of the JM approach with a recency-weighted cumulative effect, we have calculated the martingale-based residuals for both PO and NB longitudinal sub-models. Figure 5 shows the corresponding plots of the martingale residuals versus the fitted values for the emergency claims per year. In general, these residuals are skewed towards negative values, and for this reason it is very useful to superimpose a loess function (solid line) to better assess the shape of the relationship between the residuals and the fitted values. In the JM with the PO longitudinal sub-model, the loess trend deviates from zero as the fitted value increases. By contrast, in the plot for the JM with NB longitudinal response, the loess curve shows almost no evidence of a trend across all the fitted values.

5. Individualized survival predictions

Using the Bayesian joint framework, personalized and dynamically-updated survival predictions can be obtained by considering each subject-specific longitudinal profile (Proust-Lima and Taylor, 2009; Rizopoulos, 2012; Serrat et al., 2015). Let us consider a new subject, denoted by $k = i+1$, not included in the original dataset but sampled from the target population. If emergency claims per year are recorded until time t, we implicitly know that this new subject is alive at least until t, thus providing a historical set of observed measurements, $Y_k(s) = \{y_k(s_{kj}),\ \tau_k \leq s_{kj} \leq t,\ j = 1,\ldots,n_k\}$, as well as a specific value for gender factor w_{gk}. From this information, we want to predict the conditional subject-specific survival probabilities at any future time $u > t$, given survival up to t: $\tilde{p}_k(u\,|\,s) = p_t\left(T_k^* \geq u\,|\,T_k^* > s, Y_k(s), w_{gk}, D_n\right)$. This prognosis task can be carried out quite straightforwardly by adopting a Bayesian strategy. Let $\boldsymbol{\Omega} = (\boldsymbol{\theta}, \boldsymbol{b}_k)$ denote the full vector of uncertainties in the joint model and the random effects of the new subject. Then, the the conditional survival probability can be estimated from the posterior predictive distribution of the observed data:

$$
\begin{aligned}
&\tilde{p}_k(u\,|\,s) = \\
&= \iint_{\boldsymbol{\Omega}} p_t\left(T_k^* \geq u\,|\,T_k^* > s, Y_k(s), w_{gk}, \boldsymbol{b}_k, \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\,|\,D_n\right) \mathrm{d}\boldsymbol{b}_k\, \mathrm{d}\boldsymbol{\theta} \qquad (14)\\
&= \iint_{\boldsymbol{\Omega}} p_t\left(T_k^* \geq u\,|\,T_k^* > s, \boldsymbol{b}_k, \boldsymbol{\theta}\right) p_b\left(\boldsymbol{b}_k\,|\,T_k^* > s, Y_k(s), w_{gk}, \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\,|\,D_n\right) \mathrm{d}\boldsymbol{b}_k\, \mathrm{d}\boldsymbol{\theta}\\
&= \iint_{\boldsymbol{\Omega}} \frac{\Pr\left(T_k^* \geq u\,|\,M_k(u), \boldsymbol{b}_k, \boldsymbol{\theta}\right)}{\Pr\left(T_k^* > s\,|\,M_k(s), \boldsymbol{b}_k, \boldsymbol{\theta}\right)} p_b\left(\boldsymbol{b}_k\,|\,T_k^* > s, Y_k(s), w_{gk}, \boldsymbol{\theta}\right) \pi\left(\boldsymbol{\theta}\,|\,D_n\right) \mathrm{d}\boldsymbol{b}_k\, \mathrm{d}\boldsymbol{\theta}.
\end{aligned}
$$

Using the previous result and the MCMC sampling information from $\pi(\boldsymbol{\theta}, \boldsymbol{b}_i \,|\, D_n)$ for a particular JM fit (we assume that the inclusion of a new subject does not entail the updating of $\bar{\boldsymbol{\theta}}$), a simulation scheme can be applied to obtain a Monte Carlo estimate of $\tilde{p}_k(u \,|\, s)$. By way of an example, let us consider a male and female policyholders, both aged 70 years upon entering the study ($\tau_k = 5$), and not included in the original dataset. The same number of emergency claims per year is assumed for them during the next decade, those being observed annually between the ages of 70 and 80, providing $\{Y_k(s), s = 5, \ldots, 15\}$. Moreover, we assume a NB counting sequence within the JM approach with a recency-weighted cumulative effect. We first focus on estimating the survival probability of both subjects at 90 years of age, conditioned on their being alive at s, $\tilde{p}_k(u = 25 \,|\, s)$. The results are obtained by adapting the code of `survfitJM(·)` function from the `JMbayes` package, and they show how the Monte Carlo estimates update dynamically as new longitudinal information is considered (Table 5). Time-dynamic updating of this kind emphasizes the need for a well-characterized follow-up for each policyholder when we aim for personalized decisions and an accurate prediction of the insurance capital needed to cover the corresponding health insurance plan.

Table 5: *Time-dynamic probabilities of being alive at 90 years for a man and a woman with the same longitudinal information collected between the ages of 70 and 80. The results are estimated from the JM with a recency-weighted cumulative effect parameterization and a NB longitudinal sub-model.*

Age (yr.)	Emergency claims per year, $y_k(s)$	Man's survival at 90 yr.			Woman's survival at 90 yr.		
		Mean	$q_{2.5\%}$	$q_{97.5\%}$	Mean	$q_{2.5\%}$	$q_{97.5\%}$
70	0	0.783	0.478	0.864	0.796	0.509	0.871
71	0	0.804	0.601	0.868	0.817	0.628	0.874
72	1	0.790	0.550	0.863	0.803	0.563	0.872
73	0	0.806	0.635	0.868	0.818	0.647	0.874
74	2	0.776	0.556	0.854	0.790	0.585	0.863
75	0	0.793	0.616	0.861	0.807	0.632	0.871
76	0	0.807	0.660	0.867	0.822	0.682	0.875
77	8	0.671	0.359	0.812	0.692	0.367	0.826
78	1	0.690	0.404	0.816	0.708	0.410	0.829
79	2	0.688	0.406	0.812	0.706	0.414	0.827
80	0	0.719	0.479	0.825	0.736	0.501	0.839

We conclude that there is an increasing probability of being alive at the age of 90 when no emergency claims are observed, whereas this probability decreases sharply when a large number of emergency claims are annually reported. The survival probabilities for the female are higher than those for the male policyholder, since the gender regression coefficient indicates that *ceteris paribus* males have a higher mortality risk than females. Hence, by the age of 80, the survival estimate at 90 years of age for the male policyholder is $\tilde{p}_{k,m}(u = 25 \,|\, s = 15) = 0.719$, whereas a woman with the same demand presents an estimate of $\tilde{p}_{k,w}(u = 25 \,|\, s = 15) = 0.736$.

Table 6: *Time-dynamic probabilities of being alive above 80 years for a man and a woman with the same longitudinal information collected between the ages of 70 and 80. The results are estimated from the JM with a recency-weighted cumulative effect parameterization and a NB longitudinal sub-model.*

Age (years)	Man's survival above 80 yr.			Woman's survival above 80 yr.		
	Mean	$q_{2.5\%}$	$q_{97.5\%}$	Mean	$q_{2.5\%}$	$q_{97.5\%}$
80	1.000	1.000	1.000	1.000	1.000	1.000
82	0.972	0.945	0.983	0.974	0.945	0.985
84	0.934	0.871	0.960	0.939	0.875	0.964
86	0.882	0.773	0.929	0.890	0.779	0.935
88	0.812	0.640	0.884	0.824	0.655	0.896
90	0.719	0.479	0.825	0.736	0.501	0.839

If we know for certain that both subjects from the previous example remain alive when they are 80 years old, then we can also assess their future survival from the information contained in our dataset of policyholders above the age of 80. Table 6 provides the survival estimates of these two subjects. Recall that logically the last row in this table provides the same results as those in Table 5, since both survival estimates at the age of 90 are performed under the same assumptions.

6. Conclusions

Health insurance companies have access to a valuable source of information for use in follow-up studies, as they keep records of the medical claims made by each of their policyholders, in what is a discrete counting process. In this article, we have assessed the degree of relationship between an elderly policyholder's annual demand of medical emergency claims (as a longitudinal discrete response) and his or her time until death. We defined elderly people as those with a chronological age of 65 years or above, so the event times are left-truncated for all subjects whose observation time starts after this age threshold.

A correct statistical analysis of the association between the longitudinal and time-to-event outcomes entails a joint modelling approach. The longitudinal analysis was carried out using either a PO or a NB mixed model, whereas for the survival analysis, a time-dependent Cox model was used. The JM for annual rate counts and delayed entries was fitted under the Bayesian paradigm via JAGS software, entailing the challenging task of applying it to a large dataset. First, we examined the influence of the current longitudinal outcome on mortality risk. Then, in a second stage, we considered the effect of the recency-weighted area under the longitudinal profile on survival. In both cases, the results show that relatively high cumulative demand for ambulance services, hospitalizations, and non-routine visits is positively related to a deterioration in the subject's health status and, consequently, it entails higher mortality risk (i.e., lower survival probabilities). The most interesting conclusion is that the most recent critical medical

demand has the greatest impact on the current survival. This is what the JM is able to capture by means of the functional form which relates the recency-weighted area under the expected longitudinal profile with the time-to-event outcome, this approach being preferable to the one which only takes into account the effect of the current expected longitudinal value. Moreover, the results confirm the adequacy of assuming a NB distribution in the longitudinal sub-model as a first step to account for overdispersion in the longitudinal response. However, further extensions in the longitudinal part can be considered to specifically deal with zero inflation, such as different versions of zero-inflated and Hurdle models. To conclude, subject-specific survival predictions have been obtained as an example of the enormous potential of joint analysis as a predictive tool.

Acknowledgments

The authors thank the two anonymous reviewers of this article, whose valuable suggestions greatly contributed to the improvement of the manuscript.

References

Abrahamowicz, M., Beauchamp, M.-E. and Sylvestre, M.-P. (2011). Comparison of alternative models for linking drug exposure with adverse effects. *Statistics in Medicine*, 31, 1014–1030.

Barlow, W. and Prentice, R. (1988). Residuals for relative risk regression. *Biometrika*, 75, 65–74.

Boucher, J.-P., Denuit, M. and Guillén, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. *Variance*, 2, 135–162.

Cameron, C. and Trivedi, K. (1998). *Regression Analysis of Count Data*. Number 30. Cambridge University Press. Cambridge, England.

Charpentier, A. (2015). *Computational Actuarial Science with R*. Boca Raton, Florida: Chapman and Hall/CRC The R Series.

Chen, B., Covinsky, K., Stijacic, C., Adler, N. and Williams, B. (2012). Subjective social status and functional decline in older adults. *Journal of General Internal Medicine*, 27, 693–699.

Dow, W., Schoeni, R., Adler, N. and Stewart, J. (2010). Evaluating the evidence base: policies and interventions to address socioeconomic status gradients in health. *Annals of the New York Academy of Sciences*, 1186, 240–251.

Dunn, K. and Smyth, G. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236–244.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121.

Eilers P., Marx, B. and Durbán, M. (2015). Twenty years of P-splines. *SORT*, 39, 149–186.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: theory. *Econometrica*, 52, 681–700.

Greene, W. (2008). Functional forms for the negative binomial model for count data. *Economics Letters*, 99, 585–590.

Harrison, X. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ 2*, e616.

Hartig, F. (2017). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models*. R package version 0.1.5.

Hilbe, J. (2011). *Negative Binomial Regression*, 2nd Edition. Cambridge: Cambridge University Press.

Hinde, J. and Demétrio, C. (1998). Overdispersion: models and estimation. *Computational Statistics and Data Analysis*, 27, 151–170.

Ismail, N. and Jemain, A. (2007). Handling overdispersion with negative binomial and generalized Poisson regression models. In *Casualty Actuarial Society Forum*, Winter 2007, pp. 103–158.

Ivanova, A., Molenberghs, G. and Verbeke, G. (2016). Mixed models approaches for joint modelling of different types of responses. *Journal of Biopharmaceutical Statistics*, 26, 601–618.

Klein, J. and Moeschberger, M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

Lamarca, R., Alonso, J., Gómez, G. and Muñoz, A. (1998). Left-truncated data with age as time scale: an alternative for survival analysis in the elderly population. *Journal of Gerontology: Medical Sciences*, 53, M337–M343.

Lawless, J. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics*, 15, 209–225.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.

Murawska, M., Rizopoulos, D. and Lessaffre, E. (2012). A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *Journal of Probability and Statistics*, vol. in press, 2012, 1–18.

Piulachs, X., Alemany, R., Guillén, M. and Serrat, C. (2015). Joint modelling of health care usage and longevity uncertainty for an insurance portfolio. In *Scientific Methods for the Treatment of Uncertainty in Social Sciences*, pp. 289–297. Springer International Publishing.

Plummer, M. (2003). JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Technische Universität Wien, Vienna, Austria.

Proust-Lima, C. and Taylor, J. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modelling approach. *Biostatistics*, 10, 535–549.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*. Chapman and Hall/CRC Biostatistic Series. Boca Raton, Florida.

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *Journal of Statistical Software*, 72, 1–46.

Rizopoulos, D. and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30, 1366–1380.

Serrat, C., Rue, M., Armero, C., Piulachs, X., Perpiñán, H., Forte, A., Páez, A. and Gómez, G. (2015). Frequentist and Bayesian approaches for a joint model for prostate cancer risk and longitudinal prostate-specific antigen data. *Journal of Applied Statistics*, 42, 1223–1239.

Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64, 583–639.

Thiébaut, A. and Bénichou, J. (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine*, 23, 3803–3820.

Tsiatis, A. and Davidian, M. (2004). Joint Modelling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, 14, 809–834.

Uzunogullari, U. and Wang, J. (1992). A comparison of hazard rate estimators for left-truncated and right-censored data. *Biometrika*, 79, 297–310.

Viviani, S., Alfó, M. and Rizopoulos, D. (2012). Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing*, 24, 417–427.

Yu, M., Taylor, J. and Sandler, H. (2008). Individual prediction in prostate cancer studies using a joint longitudinal survival-cure model. *Journal of the American Statistical Association*, 103, 178–187.

Zuur, A., Ieno, E., Walker, N., Saveliev, A. and Smith, G. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer.